Assessing the robustness of recommendations made in a guideline on specialist neonatal respiratory care in babies born preterm with threshold analysis

David M. Phillippo[1,2], Caitlin Daly[1,2], Sofia Dias[1,2,3]

[1]Bristol Medical School (Population Health Sciences), University of Bristol, UK

[2]NICE Guidelines Technical Support Unit, University of Bristol, UK

[3]Centre for Reviews and Dissemination, University of York, UK

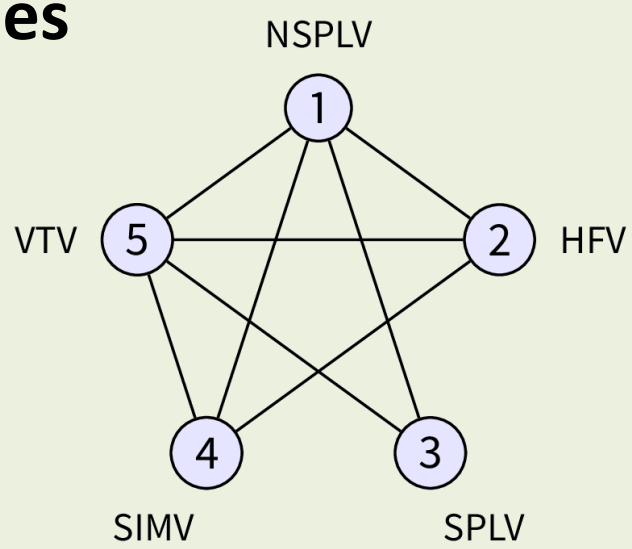
NICE Clinical Guideline: Invasive ventilation for preterm babies

Five interventions

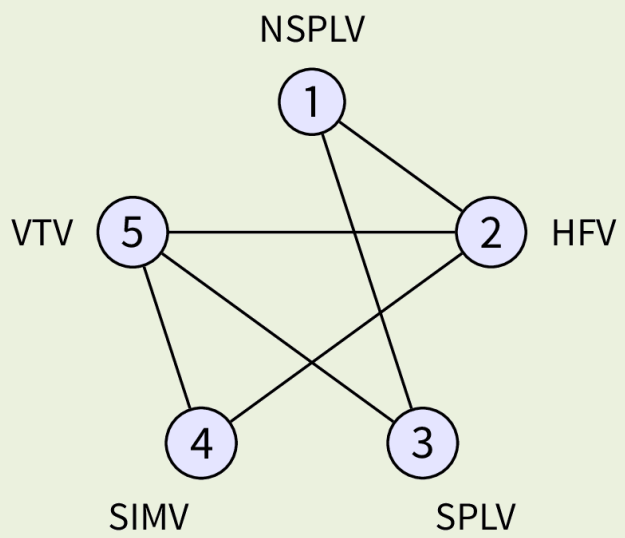
- Non-synchronised pressure limited (NSPLV)
- High frequency (HFV)
- Synchronised intermittent mandatory (SIMV)
- Synchronised pressure limited (SPLV)
- Volume targeted (VTV)

Two outcomes

Mortality



Bronchopulmonary Dysplasia (BPD)



University of BRISTOL

Network Meta-Analysis (NMA)

- Combines evidence on multiple treatments from several studies
- Arranges treatments on a network structure joined by study evidence
- Provides a coherent set of treatment effect estimates
- Is routinely used to inform clinical guideline recommendations, technology appraisals

Invasive ventilation: Initial recommendations

Invasive ventilation techniques in the neonatal unit

For preterm babies who need invasive ventilation, **use volume-targeted ventilation (VTV)** as the primary mode of respiratory support. If VTV is not effective, **consider high-frequency oscillatory ventilation (HFV)**.

Do not use synchronised pressure-limited ventilation (SPLV) such as…

(NICE guideline NG124)

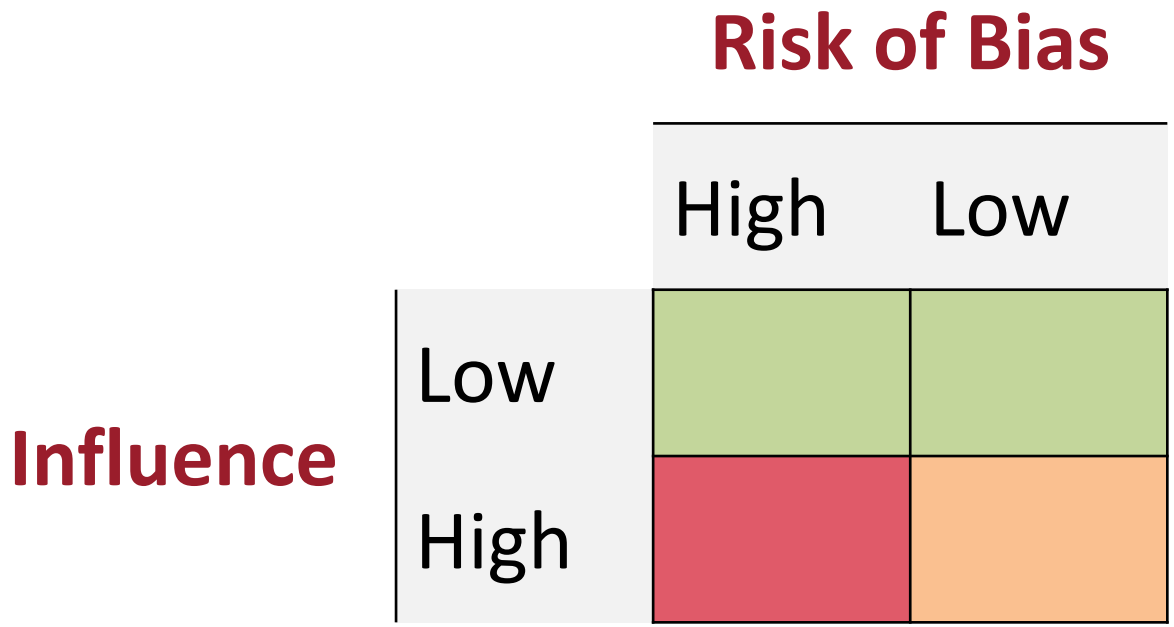
Invasive ventilation: Key concerns for decision makers

- High risk of bias due lack of blinding and treatment switching in all studies
- Concern over strength of “do-not-do” SPLV recommendation based on mortality

Motivation for threshold analysis

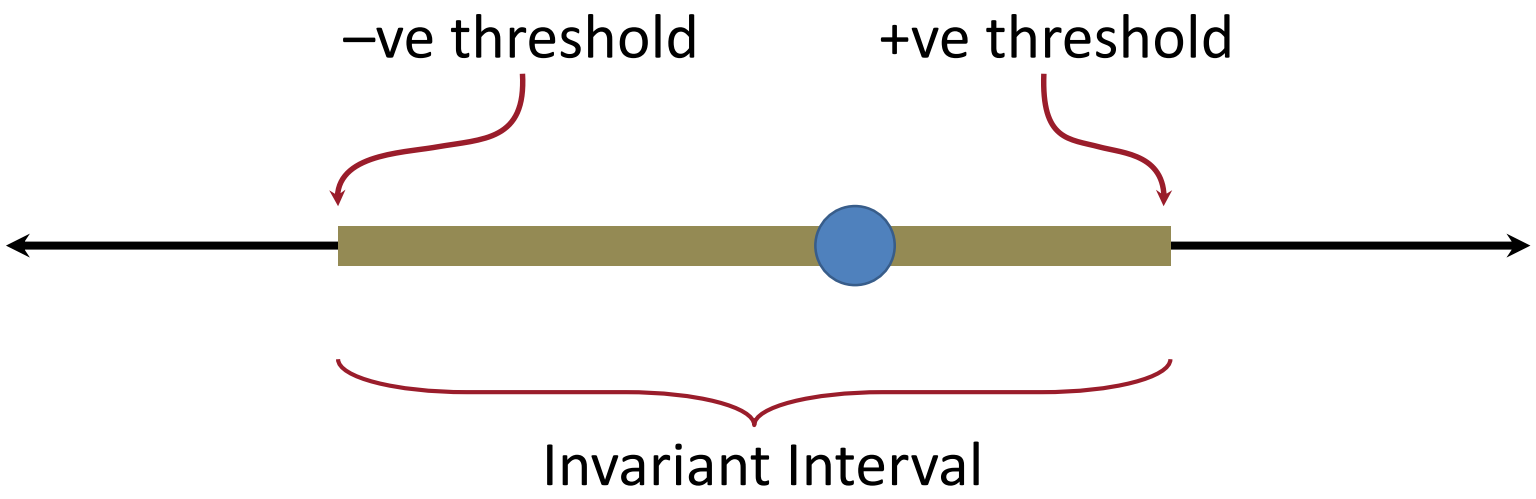
How robust are the recommendations based on NMA?

- Risk of Bias (or evidence quality) is only part of the story…



Threshold Analysis

Create an *invariant interval* for a data point:



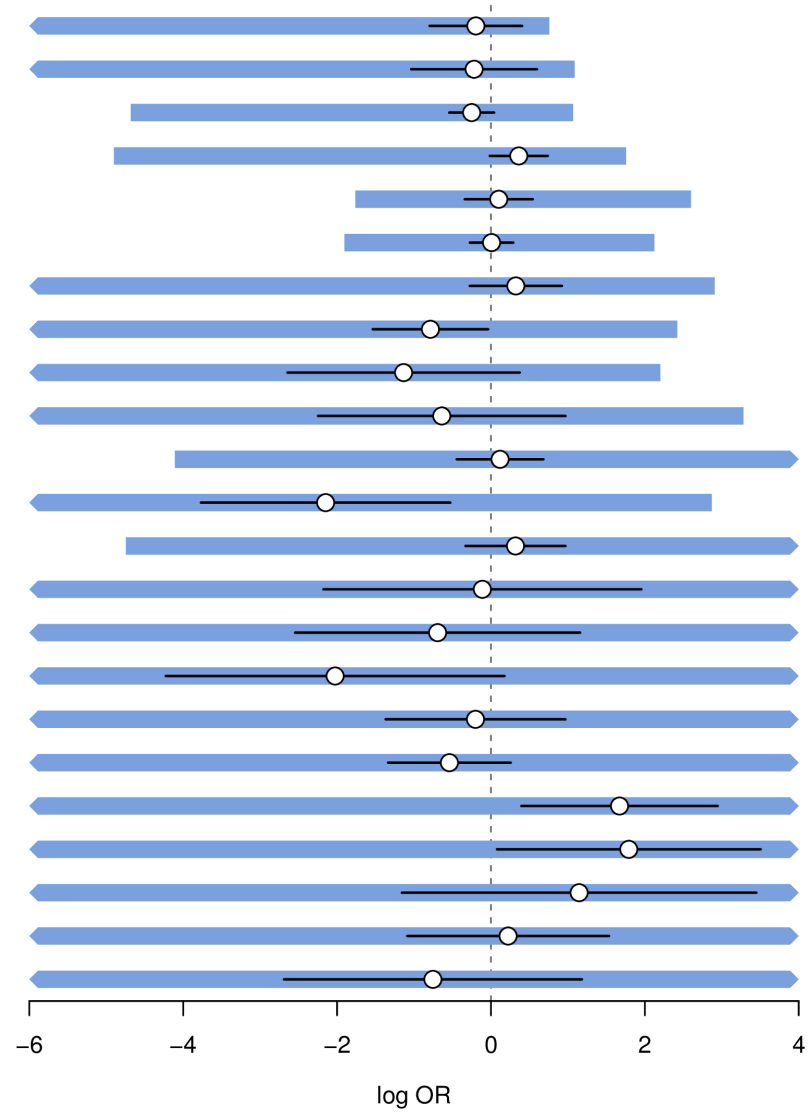
How much would the evidence have to change before we reach a new recommendation?

Positive recommendation for VTV

Threshold results for BPD and mortality

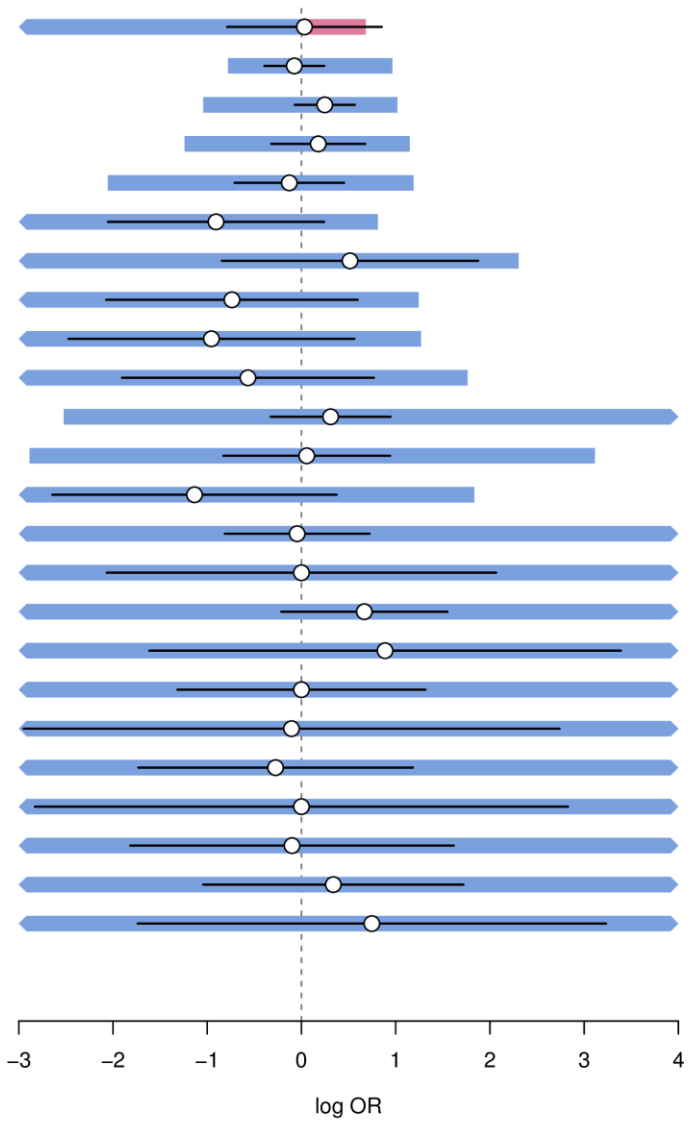
BPD – best ranked intervention (VTV, 5)

Study (contrast)	log OR	95% Confidence Interval	Invariant Interval			Sel	Perf	Det	Att	Rep	Oth
D'Angio 2005 (5 vs. 4)	−0.196	(−0.798, 0.406)	–	(NT, 0.759)	2	U	H	L	L	U	H
Singh 2006 (5 vs. 3)	−0.219	(−1.04, 0.599)	–	(NT, 1.09)	3	U	H	L	L	U	H
Baumer 2000 (3 vs. 1)	−0.250	(−0.542, 0.0420)	3	(−4.68, 1.07)	1	U	H	L	L	U	L
Courtney 2002 (4 vs. 2)	0.362	(−0.0152, 0.739)	4	(−4.90, 1.76)	2	U	H	L	U	U	H
Bereseford 2000 (3 vs. 1)	0.102	(−0.340, 0.544)	3	(−1.76, 2.60)	2	L	H	L	L	U	L
Johnson 2002 (2 vs. 1)	0.0079	(−0.274, 0.290)	2	(−1.90, 2.12)	1	U	H	L	L	L	H
Moriette 2001 (4 vs. 2)	0.324	(−0.275, 0.922)	4	(−9.18, 2.91)	2	L	H	L	L	L	H
Gerstmann 1996 (2 vs. 1)	−0.786	(−1.54, −0.0361)	2	(−8.43, 2.42)	3	H	H	L	L	U	H
Dunman 2012 (5 vs. 3)	−1.13	(−2.64, 0.374)	–	(NT, 2.20)	3	U	H	L	L	U	L
Lista 2004 (5 vs. 3)	−0.639	(−2.25, 0.969)	–	(NT, 3.28)	3	U	H	L	L	U	H
Thome 1999 (2 vs. 1)	0.119	(−0.445, 0.682)	2	(−4.11, 5.10)	1	U	H	L	L	U	L
Guven 2013 (5 vs. 4)	−2.15	(−3.77, −0.528)	–	(NT, 2.87)	2	U	H	L	L	U	H
Van reempts 2003 (2 vs. 1)	0.319	(−0.330, 0.969)	2	(−4.74, 7.32)	1	U	H	L	L	U	H
Lista 2008 (5 vs. 2)	−0.111	(−2.18, 1.96)	–	(NT, 5.54)	2	U	H	L	L	U	L
Nafday 2005 (5 vs. 4)	−0.693	(−2.55, 1.16)	–	(NT, 5.57)	2	U	H	L	L	U	L
Sinha 1997 (5 vs. 3)	−2.03	(−4.23, 0.176)	–	(NT, 4.65)	3	U	H	L	L	U	H
Craft 2003 (4 vs. 2)	−0.201	(−1.37, 0.969)	4	(−36.0, 7.03)	2	L	H	L	H	U	L
Reyes 2006 (4 vs. 4)	−0.540	(−1.34, 0.257)	–	(NT, 8.35)	3	U	H	L	L	U	L
Durand 2001 (4 vs. 2)	1.67	(0.395, 2.95)	4	(−14.4, 11.2)	2	U	H	L	L	U	H
Vento 2005 (4 vs. 2)	1.79	(0.0787, 3.50)	4	(−28.2, 17.7)	2	U	H	L	L	U	L
Salvo 2012 (4 vs. 2)	1.15	(−1.16, 3.45)	4	(−64.9, 27.4)	2	L	H	L	L	U	L
Unal 2017 (5 vs. 5)	0.223	(−1.09, 1.53)	3	(−100, NT)	–	U	H	L	L	U	L
Ozdemir 2017 (5 vs. 5)	−0.754	(−2.69, 1.18)	–	(NT, 111)	3	U	H	L	L	U	H



Mortality – best ranked intervention (VTV, 5)

Study (contrast)	log OR	95% Confidence Interval		Invariant Interval		Sel	Perf	Det	Att	Rep	Oth
D'Angio 2005 (5 vs. 4)	0.0321	(−0.788, 0.853)	–	(NT, 0.683)	4	U	H	L	L	U	H
Johnson 2002 (2 vs. 1)	−0.0758	(−0.394, 0.242)	2	(−0.780, 0.964)	1	U	H	L	L	L	H
Baumer 2000 (3 vs. 1)	0.247	(−0.0719, 0.567)	3	(−1.04, 1.02)	2	U	H	L	L	U	L
Courtney 2002 (4 vs. 2)	0.178	(−0.320, 0.677)	4	(−1.24, 1.15)	2	U	H	L	U	U	H
Moriette 2001 (4 vs. 2)	−0.129	(−0.710, 0.452)	4	(−2.06, 1.19)	2	L	H	L	L	L	H
Singh 2006 (5 vs. 3)	−0.907	(−2.05, 0.241)	–	(NT, 0.811)	2	U	H	L	L	U	H
Piotrowski 2007 (5 vs. 4)	0.515	(−0.845, 1.88)	–	(NT, 2.30)	4	U	H	L	L	U	H
Piotrowski 1997 (5 vs. 1)	−0.738	(−2.07, 0.597)	–	(NT, 1.24)	2	U	H	L	L	U	H
Guven 2013 (5 vs. 4)	−0.956	(−2.47, 0.561)	–	(NT, 1.27)	4	U	H	L	L	U	H
Lista 2004 (5 vs. 3)	−0.568	(−1.91, 0.769)	–	(NT, 1.76)	2	U	H	L	L	U	H
Van reempts 2003 (2 vs. 1)	0.309	(−0.328, 0.947)	2	(−2.52, 4.49)	1	U	H	L	L	U	H
Bernstein 1996 (4 vs. 1)	0.0561	(−0.829, 0.941)	4	(−2.89, 3.11)	1	L	H	L	L	U	L
Dunman 2012 (5 vs. 3)	−1.13	(−2.64, 0.374)	–	(NT, 1.83)	2	U	H	L	L	U	L
Thome 1999 (2 vs. 1)	−0.0455	(−0.814, 0.723)	2	(−4.17, 6.04)	1	U	H	L	L	U	L
Chowdhury 2013 (5 vs. 4)	0	(−2.07, 2.07)	–	(NT, 4.13)	4	U	H	L	L	U	H
Bereseford 2000 (3 vs. 1)	0.667	(−0.215, 1.55)	3	(−9.19, 6.56)	2	L	H	L	L	U	L
Nafday 2005 (5 vs. 4)	0.887	(−1.62, 3.39)	–	(NT, 6.94)	4	U	H	L	L	U	L
Salvo 2012 (4 vs. 2)	0	(−1.32, 1.32)	4	(−9.89, 6.77)	2	L	H	L	L	U	L
Lista 2008 (5 vs. 2)	−0.105	(−2.95, 2.74)	–	(NT, 7.65)	2	U	H	L	L	U	L
Durand 2001 (4 vs. 2)	−0.274	(−1.73, 1.18)	4	(−12.4, 8.02)	2	U	H	L	L	U	H
Sinha 1997 (5 vs. 3)	0	(−2.83, 2.83)	–	(NT, 10.4)	2	U	H	L	L	U	H
Craft 2003 (4 vs. 2)	−0.100	(−1.82, 1.62)	4	(−16.9, 11.4)	2	L	H	L	H	U	L
Rettwitz−volk 1998 (2 vs. 1)	0.338	(−1.04, 1.72)	2	(−13.0, 20.0)	1	U	H	L	L	U	H
Vento 2005 (4 vs. 2)	0.747	(−1.74, 3.23)	4	(−34.5, 24.9)	2	U	H	L	L	U	L
Gerstmann 1996 (2 vs. 1)	−Inf	(–, –)	–	(NT, NT)	–	H	H	L	L	U	H
Ogawa 1993 (2 vs. 1)	−Inf	(–, –)	–	(NT, NT)	–	U	H	L	L	U	H



○ log OR — 95% Confidence Interval ▬ Invariant Interval

Positive recommendation for VTV

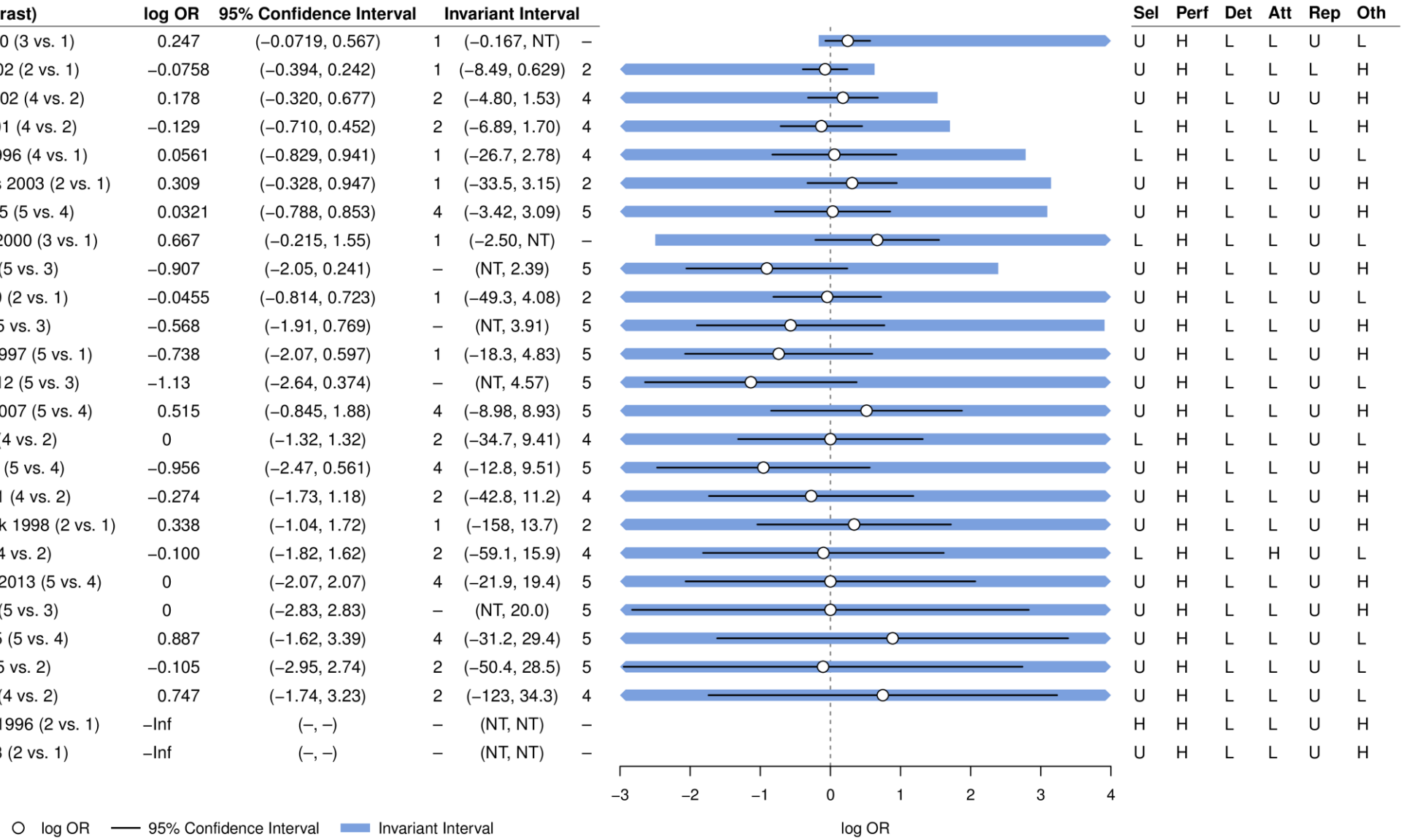
- Robust on BPD outcome, despite high risk of bias
- Sensitive to the level of uncertainty in a single study (D'Angio 2005) for mortality outcome
  - Could lead to SIMV being ranked best for mortality
  - No "significant" difference between top two treatments for mortality
    - log odds ratio for VTV vs. SIMV is –0.21 (95% CrI: –0.67, 0.25)

Do-not-do recommendation for SPLV

Threshold results for mortality

Mortality – worst ranked intervention (SPLV, 3)

Study (contrast)	log OR	95% Confidence Interval		Invariant Interval		Sel	Perf	Det	Att	Rep	Oth
Baumer 2000 (3 vs. 1)	0.247	(–0.0719, 0.567)	1	(–0.167, NT)	–	U	H	L	L	U	L
Johnson 2002 (2 vs. 1)	–0.0758	(–0.394, 0.242)	1	(–8.49, 0.629)	2	U	H	L	L	L	H
Courtney 2002 (4 vs. 2)	0.178	(–0.320, 0.677)	2	(–4.80, 1.53)	4	U	H	L	U	U	H
Moriette 2001 (4 vs. 2)	–0.129	(–0.710, 0.452)	2	(–6.89, 1.70)	4	L	H	L	L	L	H
Bernstein 1996 (4 vs. 1)	0.0561	(–0.829, 0.941)	1	(–26.7, 2.78)	4	L	H	L	L	U	L
Van reempts 2003 (2 vs. 1)	0.309	(–0.328, 0.947)	1	(–33.5, 3.15)	2	U	H	L	L	U	H
D'Angio 2005 (5 vs. 4)	0.0321	(–0.788, 0.853)	4	(–3.42, 3.09)	5	U	H	L	L	U	H
Bereseford 2000 (3 vs. 1)	0.667	(–0.215, 1.55)	1	(–2.50, NT)	–	L	H	L	L	U	L
Singh 2006 (5 vs. 3)	–0.907	(–2.05, 0.241)	–	(NT, 2.39)	5	U	H	L	L	U	H
Thome 1999 (2 vs. 1)	–0.0455	(–0.814, 0.723)	1	(–49.3, 4.08)	2	U	H	L	L	U	L
Lista 2004 (5 vs. 3)	–0.568	(–1.91, 0.769)	–	(NT, 3.91)	5	U	H	L	L	U	H
Piotrowski 1997 (5 vs. 1)	–0.738	(–2.07, 0.597)	1	(–18.3, 4.83)	5	U	H	L	L	U	H
Dunman 2012 (5 vs. 3)	–1.13	(–2.64, 0.374)	–	(NT, 4.57)	5	U	H	L	L	U	L
Piotrowski 2007 (5 vs. 4)	0.515	(–0.845, 1.88)	4	(–8.98, 8.93)	5	U	H	L	L	U	H
Salvo 2012 (4 vs. 2)	0	(–1.32, 1.32)	2	(–34.7, 9.41)	4	L	H	L	L	U	L
Guven 2013 (5 vs. 4)	–0.956	(–2.47, 0.561)	4	(–12.8, 9.51)	5	U	H	L	L	U	H
Durand 2001 (4 vs. 2)	–0.274	(–1.73, 1.18)	2	(–42.8, 11.2)	4	U	H	L	L	U	H
Rettwitz–volk 1998 (2 vs. 1)	0.338	(–1.04, 1.72)	1	(–158, 13.7)	2	U	H	L	L	U	H
Craft 2003 (4 vs. 2)	–0.100	(–1.82, 1.62)	2	(–59.1, 15.9)	4	L	H	L	H	U	L
Chowdhury 2013 (5 vs. 4)	0	(–2.07, 2.07)	4	(–21.9, 19.4)	5	U	H	L	L	U	H
Sinha 1997 (5 vs. 3)	0	(–2.83, 2.83)	–	(NT, 20.0)	5	U	H	L	L	U	H
Nafday 2005 (5 vs. 4)	0.887	(–1.62, 3.39)	4	(–31.2, 29.4)	5	U	H	L	L	U	L
Lista 2008 (5 vs. 2)	–0.105	(–2.95, 2.74)	2	(–50.4, 28.5)	5	U	H	L	L	U	L
Vento 2005 (4 vs. 2)	0.747	(–1.74, 3.23)	2	(–123, 34.3)	4	U	H	L	L	U	L
Gerstmann 1996 (2 vs. 1)	–Inf	(–, –)	–	(NT, NT)	–	H	H	L	L	U	H
Ogawa 1993 (2 vs. 1)	–Inf	(–, –)	–	(NT, NT)	–	U	H	L	L	U	H



Do-not-do recommendation for SPLV

- Worst-place ranking of SPLV for mortality was robust, despite high risk of bias

Impact on decision-making

Positive recommendation (VTV)

- First-place ranking for BPD was robust
  - Placated concerns over potential biases in the evidence
- First-place ranking for mortality was sensitive to imprecision from a single study
  - Committee considered this when formulating the positive recommendation

Do-not-do recommendation (SPLV)

- Last-place ranking for mortality was robust
  - Provided reassurance to the committee for this do-not-do recommendation

Conclusions

- Evidence quality and risk of bias are not sufficient to assess robustness of decisions
- Threshold analysis provides insight into the effects of changes in the evidence on treatment decisions
  - We can have more confidence in recommendations where thresholds are large
  - We can focus attention on the quality of decision-sensitive evidence
- More complex analyses can investigate specific concerns in the evidence
- Can be used with a range of decision rules or for decisions based on cost-effectiveness

Thank You

The views expressed in this presentation are those of the authors and not necessarily those of the funders.

This work was supported by the Centre for Clinical Practice, NICE, with funding to the Clinical Guidelines Technical Support Unit, University of Bristol.

This work was undertaken with the support of the MRC grant MR/P015298/1.

Phillippo DM et al. (2018). *Sensitivity of treatment recommendations to bias in network meta-analysis*. J Royal Stat Soc A, 181:843-867. doi:10.1111/rssa.12341

Phillippo DM et al. (2019). *Confidence in recommendations based on Network Meta-Analysis: threshold analysis as an alternative to GRADE NMA in guideline development.* Ann Intern Med, 170(8):538-546. doi:10.7326/M18-3542

R package *nmathresh* package available on CRAN.

Final guideline available from **nice.org.uk/guidance/ng124**.